



# UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE  
United States Patent and Trademark Office  
Address: COMMISSIONER FOR PATENTS  
P.O. Box 1450  
Alexandria, Virginia 22313-1450  
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/672,248	09/29/2003	Monika Henzinger	0026-0043	8702

44989 7590 05/18/2006

HARRITY SNYDER, LLP  
11350 Random Hills Road  
SUITE 600  
FAIRFAX, VA 22030

EXAMINER

RUTLEDGE, AMELIA L

ART UNIT PAPER NUMBER

2176

DATE MAILED: 05/18/2006

Please find below and/or attached an Office communication concerning this application or proceeding.

**Office Action Summary**

Application No.

10/672,248

Applicant(s)

HENZINGER ET AL.

Examiner

Amelia Rutledge

Art Unit

2176

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --  
**Period for Reply**

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

**Status**

- 1) ☒ Responsive to communication(s) filed on 23 February 2006.
- 2a) ☐ This action is **FINAL**. 2b) ☒ This action is non-final.
- 3) ☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

**Disposition of Claims**

- 4) ☒ Claim(s) 1-31 is/are pending in the application.
- 4a) Of the above claim(s) \_\_\_\_\_ is/are withdrawn from consideration.
- 5) ☐ Claim(s) \_\_\_\_\_ is/are allowed.
- 6) ☒ Claim(s) 1-31 is/are rejected.
- 7) ☐ Claim(s) \_\_\_\_\_ is/are objected to.
- 8) ☐ Claim(s) \_\_\_\_\_ are subject to restriction and/or election requirement.

**Application Papers**

- 9) ☐ The specification is objected to by the Examiner.
- 10) ☒ The drawing(s) filed on 23 February 2006 is/are: a) ☒ accepted or b) ☐ objected to by the Examiner.  
Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).  
Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
- 11) ☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

**Priority under 35 U.S.C. § 119**

- 12) ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
- a) ☐ All b) ☐ Some \* c) ☐ None of:
1. ☐ Certified copies of the priority documents have been received.
  2. ☐ Certified copies of the priority documents have been received in Application No. \_\_\_\_\_.
  3. ☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

\* See the attached detailed Office action for a list of the certified copies not received.

**Attachment(s)**

- |  |   |
|--|---|
| 1) <input checked="" type="checkbox"/> Notice of References Cited (PTO-892)  | 4) <input type="checkbox"/> Interview Summary (PTO-413)<br>Paper No(s)/Mail Date. _____ |
| 2) <input type="checkbox"/> Notice of Draftsperson's Patent Drawing Review (PTO-948)                                   | 5) <input type="checkbox"/> Notice of Informal Patent Application (PTO-152)             |
| 3) <input type="checkbox"/> Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08)<br>Paper No(s)/Mail Date _____ | 6) <input type="checkbox"/> Other: _____  |

### **DETAILED ACTION**

1. This action is responsive to communications: Amendment, filed 02/23/2006.
2. Claims 1-31 are pending in the case. Claims 1, 10, 15, 21, 26, and 31 are independent claims.
3. Independent claims 1 and 31 have been amended to overcome the rejections under 35 U.S.C. 112, second paragraph, therefore the rejections of 1-8 and 31 under 35 U.S.C. 112, second paragraph are withdrawn.
4. Applicants' amendments to the specification are accepted.
5. Applicant's arguments, see Remarks, p. 12-22, filed 02/23/2006, with respect to the rejections of claims 10-30 under 35 USC 103 have been fully considered and are persuasive. Therefore, the previous rejections have been withdrawn. However, upon further consideration, a new ground(s) of rejection is made over Galai in view of DaCosta.

### ***Claim Rejections - 35 USC § 103***

6. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

7. **Claims 1-31 are rejected under 35 U.S.C. 103(a) as being unpatentable over Galai et al. (hereinafter "Galai"), PCT Application filed August 2002, International Publication Number WO 03/017023 A2, published February 2003, in view of**

Art Unit: 2176

**DaCosta et al. (hereinafter "DaCosta"), U.S. Patent No. 6,665,658, issued December 2003.**

**Independent claim 1 cites:** *A method for crawling documents comprising: receiving a uniform resource locator (URL);*

*receiving at least two different copies of a document associated with the URL; and*

*determining whether a web site **corresponding to the URL** uses session identifiers based on a comparison of URLs **that are within the document and that change between the at least two different copies of the document.***

Galai teaches a method of indexing dynamic web pages for a search engine. The search engine consists of a spider and repository (p. 4, l. 3-19). Galai teaches a method for normalizing the URL of a document to index substantially similar Web pages only once (p. 20, l. 10-20). Galai teaches comparing a Web page with a second retrieved web page with reduced parameters, i.e., any divisible subunit of the URL (p. 20, l. 10-20). Galai teaches the comparison of URLs within the document where the Web page includes one or more links with the complete URL, as for a sessionID (p. 20, l. 21-p. 21, l. 9), resulting in two web pages which are similar in content but not identical. Galai teaches detecting the change between the two different copies of the document (p. 21, l. 1-8; p. 27, l. 14-p. 28, l. 21).

While Galai teaches a comparison of URLs for redundant parameters, which would include session identifiers since session identifiers are parameters within a URL, Galai does not explicitly teach that the URLs are compared for the specific purpose of

determining whether the web site uses session identifiers. However, DaCosta teaches a method for crawling documents in a dynamic website, with a database for storing and identifying session identifiers URLs, and an application program for controlling a software agent (Col. 4, l. 41-Col. 5, l. 23). DaCosta teaches the analysis of URLs and headers containing cookie data to determine if a web site uses session IDs (Col. 6, l. 21-40). It was notoriously well known in the art at the time of the invention that session data for a web site and/or document could be contained in either the URL string, or in a cookie.

Both DaCosta and Galai are directed toward methods for crawling web documents and tracking state and session information for web documents. It would have been obvious to one of ordinary skill in the art at the time of the invention to combine the method of indexing web pages for a search engine by removing redundant pages by comparing URL parameters taught by Galai, with the means of identifying session identifiers by comparing URL data and cookie data taught by DaCosta, so that Galai would have the benefit of identifying session information for a web site whether the session information were contained in the URL string or in the cookie, in order to remove redundant pages from both configurations (URL string or cookie) of dynamic web sites.

**Regarding dependent claim 2,** Galai teaches that the method of comparing URLs can be applied to any web page in a site (p. 4, l. 15-20).

**Regarding dependent claim 3,** Galai teaches a method of normalizing the URL in order to index substantially similar web pages only once (p. 20, l. 10-23); i.e.,

*comparison of the clean, or normalized URL to a set of clean URLs that represent previously crawled URLs; since Galai teaches a process for detecting redundant parameters in URLs with the same structure, executed once per URL structure, and then applied and executed for application to each URL with the same structure (p. 21, l.21-p. 22, l. 5).*

**Regarding dependent claims 4-6,** Galai teaches that the method of comparing URLs can be applied to any web page in a site (p. 4, l. 15-20). Galai teaches an automatic method of URL comparison to remove redundant parameters from pages (p. 20, l. 10-20), which would include session IDs (p. 20, l. 21-23), where the rules are determined automatically by comparing the URLs for redundancy and normalizing them.

**Regarding dependent claim 7,** Galai teaches a process for detecting redundant parameters in URLs with the same structure, executed once per URL structure, and then applied and executed for application to each URL with the same structure (p. 21, l.21-p. 22, l. 5), compare to *receiving the URL as a URL from a previously crawled web document.*

**Regarding dependent claim 8,** Galai teaches crawling the URL when the URL is determined to not already have been crawled (p. 24, l. 9-15).

**Regarding dependent claim 9,** Galai teaches comparing a portion of the URLs that change between the two copies of the document and determining similarity based on a predetermined value of the portion of the URLs that change (p. 27, l. 6- p. 28, l. 21; especially p. 28, l. 11-21), since Galai teaches automatically determining the redundant

Art Unit: 2176

parameter by URL comparison and then using that parameter as a basis of comparison to other URLs.

**Independent claim 10 cites:** *A method for identifying web sites that use session identifiers comprising: downloading at least two different copies of at least one document from a web site; extracting uniform resource locators (URLs) from the two different copies of the web document; comparing the extracted URLs of the two different copies of the document; and determining whether the web site uses session identifiers based on the comparison.*

Galai teaches a method of indexing dynamic web pages for a search engine. The search engine consists of a spider and repository (p. 4, l. 3-19). Galai teaches a method for normalizing the URL of a document to index substantially similar Web pages only once (p. 20, l. 10-20). Galai teaches comparing a Web page with a second retrieved web page with reduced parameters, i.e., any divisible subunit of the URL (p. 20, l. 10-20). Galai teaches the comparison of URLs within the document where the Web page includes one or more links with the complete URL, as for a sessionID (p. 20, l. 21-p. 21, l. 9), resulting in two web pages which are similar in content but not identical. Galai teaches detecting the change between the two different copies of the document (p. 21, l. 1-8; p. 27, l. 14-p. 28, l. 21).

While Galai teaches a comparison of URLs for redundant parameters, which would include session identifiers since session identifiers are parameters within a URL, Galai does not explicitly teach that the URLs are compared for the specific purpose of determining whether the web site uses session identifiers. However, DaCosta teaches

Art Unit: 2176

a method for crawling documents in a dynamic website, with a database for storing and identifying session identifiers URLs, and an application program for controlling a software agent (Col. 4, l. 41-Col. 5, l. 23). DaCosta teaches the analysis of URLs and headers containing cookie data to determine if a web site uses session IDs (Col. 6, l. 21-40). It was notoriously well known in the art at the time of the invention that session data for a web site and/or document could be contained in either the URL string, or in a cookie.

Both DaCosta and Galai are directed toward methods for crawling web documents and tracking state and session information for web documents. It would have been obvious to one of ordinary skill in the art at the time of the invention to combine the method of indexing web pages for a search engine by removing redundant pages by comparing URL parameters taught by Galai, with the means of identifying session identifiers by comparing URL data and cookie data taught by DaCosta, so that Galai would have the benefit of identifying session information for a web site whether the session information were contained in the URL string or in the cookie, in order to remove redundant pages from both configurations (URL string or cookie) of dynamic web sites.

**Regarding dependent claim 11**, Galai teaches comparing a portion of the URLs that change between the two copies of the document and determining similarity based on a predetermined value of the portion of the URLs that change (p. 27, l. 6- p. 28, l. 21; especially p. 28, l. 11-21), since Galai teaches automatically determining the redundant



parameter by URL comparison and then using that parameter as a basis of comparison to other URLs.

**Regarding dependent claim 12**, Galai teaches that the method of comparing URLs can be applied to any web page in a site (p. 4, l. 15-20).

**Regarding dependent claim 13**, claim 13 reflects substantially similar subject matter as claimed in dependent claim 2, and is rejected along the same rationale.

**Regarding dependent claim 14**, Galai teaches an automatic method of URL comparison to remove redundant parameters from pages (p. 20, l. 10-20), which would include session IDs (p. 20, l. 21-23), where the rules are generated automatically by the method of comparing the URLs for redundancy and normalizing them.

**Independent claim 15 cites:** *A device comprising: a spider component configured to crawl web documents associated with at least one web site; and a session identifier component configured to determine whether the web site uses session identifiers based on a comparison of a portion of uniform resource locators (URLS) that change between different copies of at least one web document downloaded from the web site.*

Galai teaches a method of indexing dynamic web pages for a search engine. The search engine consists of a spider and repository (p. 4, l. 3-19). Galai teaches a method for normalizing the URL of a document to index substantially similar Web pages only once (p. 20, l. 10-20). Galai teaches comparing a Web page with a second retrieved web page with reduced parameters, i.e., any divisible subunit of the URL (p. 20, l. 10-20). Galai teaches the comparison of URLs within the document where the

Art Unit: 2176

Web page includes one or more links with the complete URL, as for a sessionID (p. 20, l. 21-p. 21, l. 9), resulting in two web pages which are similar in content but not identical. Galai teaches detecting the change between the two different copies of the document (p. 21, l. 1-8; p. 27, l. 14-p. 28, l. 21).

While Galai teaches a comparison of URLs for redundant parameters, which would include session identifiers since session identifiers are parameters within a URL, Galai does not explicitly teach that the URLs are compared for the specific purpose of determining whether the web site uses session identifiers. However, DaCosta teaches a method for crawling documents in a dynamic website, with a database for storing and identifying session identifiers URLs, and an application program for controlling a software agent (Col. 4, l. 41-Col. 5, l. 23). DaCosta teaches the analysis of URLs and headers containing cookie data to determine if a web site uses session IDs (Col. 6, l. 21-40). It was notoriously well known in the art at the time of the invention that session data for a web site and/or document could be contained in either the URL string, or in a cookie.

Both DaCosta and Galai are directed toward methods for crawling web documents and tracking state and session information for web documents. It would have been obvious to one of ordinary skill in the art at the time of the invention to combine the method of indexing web pages for a search engine by removing redundant pages by comparing URL parameters taught by Galai, with the means of identifying session identifiers by comparing URL data and cookie data taught by DaCosta, so that Galai would have the benefit of identifying session information for a web site whether

Art Unit: 2176

the session information were contained in the URL string or in the cookie, in order to remove redundant pages from both configurations (URL string or cookie) of dynamic web sites.

**Regarding dependent claim 16-17**, Galai teaches a spider to download content from a network and a component of the autonomous software search program to extract URLs from the downloaded content (p. 26, l. 19-p. 27, l. 5) compare to *fetch component configured to download content from a network; and a content manager configured to extract URLs from the downloaded content*.

**Regarding dependent claim 18**, claim 18 reflects substantially similar subject matter as claimed in dependent claim 2, and is rejected along the same rationale.

**Regarding dependent claim 19**, Galai teaches that the method of comparing URLs can be applied to any web page in a site (p. 4, l. 15-20).

**Regarding dependent claim 20**, Galai teaches an automatic method of URL comparison to remove redundant parameters from pages (p. 20, l. 10-20), which would include session IDs (p. 20, l. 21-23), where the rules are generated automatically by comparing the URLs for redundancy and normalizing them.

**Regarding independent claim 21**, claim 21 is directed toward the device used for implementing the method as claimed in claim 10, and is rejected along the same rationale.

**Regarding dependent claims 22 and 23**, claims 22 and 23 reflect substantially similar subject matter as claimed in dependent claims 11 and 12, and are rejected along the same rationale.

**Regarding dependent claim 24**, claim 24 reflects substantially similar subject matter as claimed in dependent claim 2, and is rejected along the same rationale.

**Regarding dependent claim 25**, claim 25 reflects substantially similar subject matter as claimed in dependent claim 14, and is rejected along the same rationale.

**Regarding independent claim 26**, claim 26 is directed toward the computer-readable medium containing programming instruction for executing the method as claimed in claim 10, and is rejected along the same rationale.

**Regarding dependent claims 27 and 28**, claims 27 and 28 reflect substantially similar subject matter as claimed in dependent claims 11 and 12, and are rejected along the same rationale.

**Regarding dependent claim 29**, claim 29 reflects substantially similar subject matter as claimed in dependent claim 2, and is rejected along the same rationale.

**Regarding dependent claim 30**, claim 30 reflects substantially similar subject matter as claimed in dependent claim 14, and is rejected along the same rationale.

**Independent claim 31 cites:** *A method for crawling documents comprising: receiving a uniform resource locator (URL), and determining whether the URL is associated with a web site that uses session identifiers based on a comparison of content between different duplicate or near-duplicate copies of a document downloaded from the web site.*

Galai teaches a method of indexing dynamic web pages for a search engine. The search engine consists of a spider and repository (p. 4, l. 3-19). Galai teaches a method for normalizing the URL of a document to index substantially similar Web pages

Art Unit: 2176

only once (p. 20, l. 10-20). Galai teaches comparing a Web page with a second retrieved web page with reduced parameters, i.e., any divisible subunit of the URL (p. 20, l. 10-20). Galai teaches the comparison of URLs within the document where the Web page includes one or more links with the complete URL, as for a sessionID (p. 20, l. 21-p. 21, l. 9), resulting in two web pages which are similar in content but not identical. Galai teaches detecting the change between the two different copies of the document (p. 21, l. 1-8; p. 27, l. 14-p. 28, l. 21). Galai teaches determining whether session identifiers are used based on a comparison of content between different duplicate or near duplicate copies of a document (p. 27, l. 18-p. 28, l. 11).

While Galai teaches a comparison of URLs for redundant parameters, which would include session identifiers since session identifiers are parameters within a URL, Galai does not explicitly teach that the URLs are compared for the specific purpose of determining whether the web site uses session identifiers. However, DaCosta teaches a method for crawling documents in a dynamic website, with a database for storing and identifying session identifiers URLs, and an application program for controlling a software agent (Col. 4, l. 41-Col. 5, l. 23). DaCosta teaches the analysis of URLs and headers containing cookie data to determine if a web site uses session IDs (Col. 6, l. 21-40). It was notoriously well known in the art at the time of the invention that session data for a web site and/or document could be contained in either the URL string, or in a cookie.

Both DaCosta and Galai are directed toward methods for crawling web documents and tracking state and session information for web documents. It would

Art Unit: 2176

have been obvious to one of ordinary skill in the art at the time of the invention to combine the method of indexing web pages for a search engine by removing redundant pages by comparing URL parameters taught by Galai, with the means of identifying session identifiers by comparing URL data and cookie data taught by DaCosta, so that Galai would have the benefit of identifying session information for a web site whether the session information were contained in the URL string or in the cookie, in order to remove redundant pages from both configurations (URL string or cookie) of dynamic web sites.

### ***Response to Arguments***

8. Applicant's arguments with respect to claims 1-9 and 31 have been considered but are moot in view of the new ground(s) of rejection. The new grounds of rejection includes the Galai reference, which is being relied upon to teach the newly claimed limitations.

### ***Conclusion***

9. The prior art made of record and not relied upon is considered pertinent to applicant's disclosure.

Miller, et al., "SPHINX: A Framework for Creating Personal, Site-Specific Web Crawlers", *Proceedings of the Seventh International World Wide Web Conference*, Brisbane, Australia, April, 1998, p. 1-13.

Seda, C., "Making Dynamic and E-Commerce Sites Search Engine Friendly", *Search Engine Watch*, <http://searchenginewatch.com/searchday/article.php/2161081>, published October 29, 2002, p. 1-5.

Art Unit: 2176

Quigo, Inc., "Quigo, Inc. Unveils Deep Web Search Technologies", press release dated August 15, 2001, p. 1-2.

Search Tools Consulting, [www.searchtools.com](http://www.searchtools.com), "Generating Simple URLs for Search Engines", July 2003, p. 1-6.

Sun Microsystems, *The Java Tutorial*, including sections on "Session Tracking", "Working with URLs", and "Parsing a URL", p. 1-9.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Amelia Rutledge whose telephone number is 571-272-7508. The examiner can normally be reached on Monday - Friday 9:30 - 6:00.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Heather Herndon can be reached on 571-272-4136. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

AR

  
HEATHER R. HERNDON  
SUPERVISORY PATENT EXAMINER  
TECHNOLOGY CENTER 2100